



## Channel Selection Improves MEG-based Brain-Computer Interface

Roy, S., Rathee, D., McCreadie, K., & Prasad, G. (2019). Channel Selection Improves MEG-based Brain-Computer Interface. In *9th International IEEE EMBS Conference on Neural Engineering, NER 2019* (Vol. 2019-March, pp. 295-298). [8716948] (International IEEE/EMBS Conference on Neural Engineering). IEEE.  
<https://doi.org/10.1109/NER.2019.8716948>

[Link to publication record in Ulster University Research Portal](#)

### Published in:

9th International IEEE EMBS Conference on Neural Engineering, NER 2019

### Publication Status:

Published (in print/issue): 20/05/2019

### DOI:

[10.1109/NER.2019.8716948](https://doi.org/10.1109/NER.2019.8716948)

### Document Version

Author Accepted version

### General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [pure-support@ulster.ac.uk](mailto:pure-support@ulster.ac.uk).

# Channel Selection Improves MEG-based Brain-Computer Interface

Sujit Roy\*, Dheeraj Rathee *Student Member, IEEE*, Karl McCreadie, Girijesh Prasad *Senior Member, IEEE*

**Abstract**—This study investigates the effect of channel selection on the performance of a Magnetoencephalography (MEG)-based brain-computer interface (BCI) system in terms of classification accuracy (CA). Although many efforts are currently being undertaken to develop BCI using MEG, the major concern still is low accuracy. MEG systems involve data recording from a large number of channels which may provide a better spatio-temporal resolution for assessing brain patterns, however, a large number of channels result in a large number of features, which further make feature learning a challenging task. In this study, we evaluated the performance of two state-of-the-art channel selection methods, i.e. class-correlation (CC) and ReliefF (RF) across six binary classification tasks with a MEG dataset of 15 healthy participants. Both CC and RF methods provided a statistically significant increase in the CA (range: 20.91 - 24.22%) compared to baseline (i.e. using 204 channels) with bandpower features from the alpha (8-12 Hz) and beta frequency bands (13-30 Hz). Moreover, both methods reduce the optimum number of channels significantly (from 204 to the range of 1-22). Reducing the number of features can significantly reduce the computational cost and increase the chances of numerical stability which are key considerations in neurofeedback (online) applications.

## I. INTRODUCTION

People suffering from motor disabilities or severe neurological impairments, need a method for rehabilitation to interact efficiently with the environment. Rehabilitation of upper limb function by motor imagery (MI) practice through Brain-Computer Interface (BCI) is becoming a therapeutic alternative for stroke patients [1], [2]. For an effective rehabilitation to take place, there is a dire need on the sufferers' behalf to intensively undergo voluntary practice with the paretic limb in a focused way adhering to activities of daily living (ADL) practices [2], [3], [4], [5]. Past decades have shown the development of BCI-based rehabilitative technologies [6] that provides a novel and effective communication medium for post-stroke rehabilitation interventions. Current BCIs may involve acquisition of brain responses through various invasive and non-invasive methods such as Magnetoencephalography (MEG), Electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and Electrocorticography (ECoG) [7].

MEG-based BCIs have been the focus of several recent studies, however, the performance of these systems are still low [8]. The majority of these studies focused on either

development or implementation of efficient feature extraction algorithms rather than signal pre-processing methods such as channel selection. However, signal pre-processing can significantly improve the BCI system performance [9]. MEG systems involve acquisition of brain responses from a large number of channels (i.e. 360 for Elekta Neuromag Triux system). The MEG data from these channels may result in high features dimension. Provided with a large number of channels, the extracted features may outnumber the observation set (i.e. number of trials). It is often required to reduce the dimensionality of data without disturbing the useful features as it has been observed that the use of too many channels can negatively affect the feature separability and increase the computational cost of a BCI [10], [11]. Moreover, a high dimensional feature set may create numerical instability during the machine learning step. Thus, it is important to select the channels which contribute greatly to the discrimination of the mental tasks.

Neuro-feedback based BCI can facilitate motor improvements with moderate BCI accuracy as evidenced by Prasad et al. [2] who found promising results after examining five stroke patients. For MEG-BCI based neurorehabilitation systems, the extraction of the relevant information from brain activity is a big challenge. However, the gradiometers (204 channels) provide higher signal-to-noise ratio (SNR) as compared to magnetometers (102 channels) and are more sensitive to cortical activations, thus, it is intuitive to use MEG data from gradiometers only, particularly for an MI-BCI application. Nevertheless, the number of channels is still high enough to increase the computation irrespective of positively or negatively contributing to the features classification. Selecting the best channels not only decreases computation cost but also shows a substantial increase in accuracy. Common spatial pattern (CSP) and its extended algorithms like Sparse CSP, have been used for dimensionality reduction and show that the channels can be selected on the basis of maximal CSP vector coefficients [12], [11], [13] whilst maintaining a sufficient level of accuracy. Further, different combinations of CSP methods have been implemented in an attempt to increase the accuracy [8]. A channel selection method was presented by He et al. [14] based on the Bhattacharyya bound CSP method for classification of MI. It takes into account the Bhattacharyya bound as an index and progressively searches for optimized channel combinations. A 95% accuracy rate of the classification was later claimed with an average of 33 channels, which was higher than that of any other channel selection methods. For a review of other methods implemented in the field of EEG motor imagery, readers can refer to Turkey et. al and Lotte et. al [15], [16].

SR is supported by Department for the Economy (DfE), Northern Ireland and UKIERI DST. G.P. is supported by the NIFBM Facility project (1303/101154803) and UKIERI DST Thematic Partnership project (DST-UKIERI-2016-17-0128).

S.R., D.R., K.M. and G.P. are with the Intelligent Systems Research Centre, School of Computing & Intelligent Systems, Ulster University, Derry~Londonderry, N. Ireland, UK. Corresponding author: Sujit Roy: roy-s2@ulster.ac.uk

Whereas the previous studies involved implementation of channels selection methods on EEG data, no such work has been reported using MEG. In this study, we implemented two state-of-the-art channel selection methods i.e. class-correlation (CC) method and the ReliefF (RF) method on a mental imagery related MEG dataset and analysed their effect on the performance of binary classification tasks.

## II. MATERIAL AND METHODS

### A. MEG Dataset

The MEG dataset of fifteen healthy participants (12 males, 3 females, mean age  $29.3 \pm 5.96$  yrs., two left-handed as per self-report) was acquired with a typical BCI paradigm using the Elekta Neuromag Triux system (Fig. 1). The dataset includes 4 different mental imagery classes, namely both hand movement, both feet movement, subtraction, and word generation.

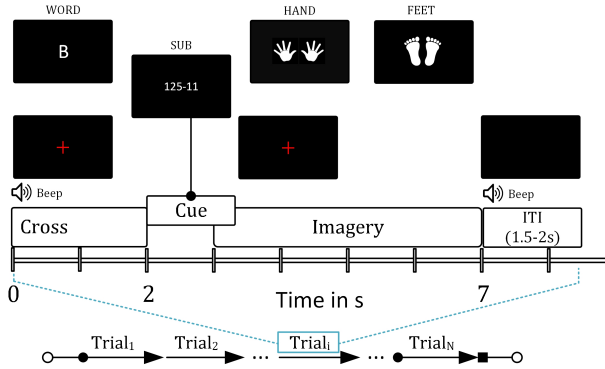


Fig. 1. Experimental paradigm for the MEG data acquisition

Fig. 1 provides the schematic representation of the signal processing pipeline implemented in this study. The raw MEG data from 204 gradiometers were band-passed in range of 8-12 Hz and 13-30 Hz and signal power values were estimated for each channel and for all trials separately. Furthermore, 10-fold cross validation classification accuracies (CAs) were estimated for six binary classification tasks i.e. hand versus feet (H-F), hand versus word (H-W), hand versus math (H-M), feet versus word (F-W), feet versus math (F-M), and word versus math (W-M) using a linear discriminant analysis (LDA) classifier. Three different experimental conditions were considered for evaluation of the effect of the channel selection. The first condition (i.e. baseline) considered features from all the channels. The second condition (CC) considered class-correlation method to select the best channels whereas the third condition (RF) involved the implementation of ReliefF method for channel selection.

### B. Class-Correlation Method

The Pearson correlation coefficient is used to determine the statistical relationship between two random variables A and B. The values of the coefficients range from -1 to 1 representing no relation to direct relation. It is a measure of the linear dependence of the two random variables [17].

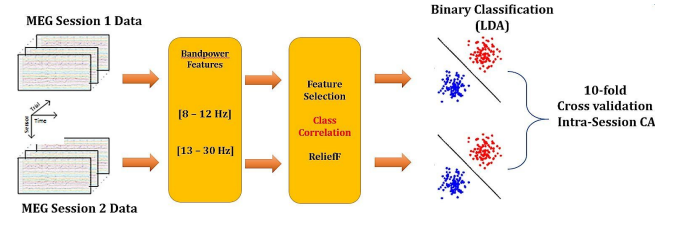


Fig. 2. Schematic diagram providing the details of signal processing pipeline.

$$\rho(A, B) = \frac{1}{N-1} \sum_{i=1}^N \left( \frac{A_i - \mu_A}{\sigma_A} \right) \left( \frac{B_i - \mu_B}{\sigma_B} \right) \quad (1)$$

where  $\mu_A$  is the mean and  $\sigma_A$  is the standard deviation of A. Similarly,  $\mu_B$  is the mean and  $\sigma_B$  is the standard deviation of B. The absolute values of correlation coefficients were estimated by creating a dummy label corresponding to the observations of each features. The dummy class label had a numeric value which indicated (l=1) for class 1 and (l=2) class 2. As per the coefficients the channels were ranked in decreasing order keeping the highest correlated channel at the top.

### C. ReliefF Method

The Relief algorithm proposed by Kira et al. [18], [19] estimates the quality of the attributes that have weights greater than the thresholds using the difference of an attribute value between a given instance and the two nearest instances (Hit and Miss). The algorithm of ReliefF is as follows:

**Data:** For each training instance a vector of attribute values and the class value

**Result:** The vector W of estimations of the qualities of attributes

Intialise the weights:  $W[A] \leftarrow 0$  ;

**for**  $i \leftarrow 1$  **to**  $m$  **do**

    randomly select an instance  $R_i$ ;

    find  $k$  nearest hits  $H_j$  ;

**for each class**  $C \neq \text{class}(R_i)$  **do**

        from class  $C$  find  $k$  nearest misses  $M_j(C)$ ;

**end**

**end**

**for**  $A \leftarrow 1$  **to**  $a$  **do**

$$W[A] = W[A] - \sum_{j=1}^k \frac{\text{diff}(A, R_i, H_j)}{m.k} + \sum_{C \neq \text{class}(R_i)} \frac{\frac{P(C)}{1-P(\text{class}(R_i))} \sum_{j=1}^k \text{diff}(A, R_i, M_j(C))}{m.k} \quad (2)$$

**end**

**Algorithm 1:** Pseudocode for ReliefF algorithm

The ReliefF algorithm is an extension of this algorithm and it differs from the Relief by selecting the number of hits and

TABLE I

COMPARISON OF THE NUMBER OF CHANNELS CONTRIBUTING TO MAXIMUM ACCURACY BETWEEN RELIEFF METHOD ( $M_1$ ) AND THE CLASS CORRELATION METHOD ( $M_2$ ) FOR THE DATASET IN ALPHA FREQ BAND.

Subjects	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15	
	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$		
H-F	14	7	8	16	14	9	1	10	12	13	8	10	15	8	6	8	7	3	11	9	12	19	3	4	8	7	6	4	8	10
H-W	14	12	18	10	13	10	12	6	13	17	6	7	14	4	13	14	6	4	11	4	10	15	9	7	9	9	16	7	13	19
H-M	12	14	14	10	19	15	10	13	15	21	17	12	8	15	12	16	8	7	13	9	15	8	7	13	10	15	8	4	6	9
F-W	11	15	15	10	17	17	13	8	9	5	17	15	15	18	15	16	12	15	10	11	21	13	5	12	16	15	15	6	11	16
F-M	15	16	13	3	19	15	18	10	20	13	11	12	9	14	13	9	9	7	8	4	13	15	5	8	4	10	17	7	17	14
W-M	11	6	13	17	12	5	12	14	13	7	3	8	9	15	13	13	15	9	22	13	9	7	7	12	12	13	3	2	16	22

TABLE II

COMPARISON OF THE NUMBER OF CHANNELS CONTRIBUTING TO MAXIMUM ACCURACY BETWEEN RELIEFF METHOD ( $M_1$ ) AND THE CLASS CORRELATION METHOD ( $M_2$ ) FOR THE DATASET IN BETA FREQ BAND.

Subjects	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15	
	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$
H-F	18	13	5	3	14	11	17	15	14	2	15	6	12	8	14	9	9	6	2	8	9	13	4	14	14	5	2	3	3	3
H-W	8	7	6	14	15	17	6	8	19	11	7	17	9	6	11	7	15	16	8	16	12	19	6	6	11	7	9	5	9	16
H-M	1	7	10	19	8	7	9	13	10	12	12	1	14	8	6	7	14	11	11	11	13	1	6	11	14	13	13	15	11	
F-W	4	11	8	7	18	11	10	8	12	9	12	12	9	16	13	10	13	11	4	9	12	17	11	16	7	9	7	6	12	15
F-M	1	3	9	13	21	19	10	9	11	7	6	5	15	6	16	5	13	4	11	7	8	9	7	3	10	7	6	8	4	11
W-M	9	5	5	16	6	13	9	4	12	17	8	8	1	8	5	7	8	17	7	8	6	10	2	4	5	10	5	10	13	15

TABLE III

MEAN CLASSIFICATION ACCURACY (SD) OBTAINED WITH THREE COMPETING METHODS FOR SIX BINARY CLASSIFICATION TASKS AND THEIR GRAND MEAN FOR  $\alpha$  BAND ( $M_1$ ) AND FOR  $\beta$  BAND ( $M_2$ ).

Methods	H-F		H-W		H-M		F-W		F-M		W-M		Mean	
	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$
Baseline	57	54.84	71.51	66.32	63.66	59.2	68.72	61.89	65.01	58.48	66	58.53	65.32	59.87
CC	(9.96)	(6.18)	(8.89)	(10.3)	(11.1)	(8.48)	(7.9)	(6.39)	(9.92)	(8.13)	(8.51)	(7.39)	(8.51)	(7.39)
	73.89	67.49	82.72	77.48	80.57	72.47	83.85	76.69	78.57	69.55	81.60	70.69	80.20	72.39
RF	(7.53)	(6.6)	(8.93)	(8.29)	(9.33)	(7.67)	(8.36)	(8.64)	(8.22)	(7.53)	(6.87)	(7.27)	(8.66)	(8.32)
	74.27	71.88	83.67	79.4	80.9	72.52	85.76	77.44	81.35	72.28	80.95	70.61	81.14	74.02
	(7.43)	(6.72)	(7.48)	(6.82)	(8.1)	(9.47)	(6.86)	(7.07)	(7.39)	(8.60)	(7.3)	(6.60)	(8.05)	(8.08)

updating the quality of attributes [20], [19]. The RF method depends on user parameter  $k$ , which declares the use of  $k$  nearest hits and  $k$  nearest misses for the weight update for each target instance. This change increases weight estimate reliability, particularly in noisy problems.

### III. RESULTS

Figs. 3 and 4 presents the bar plots of the mean classification accuracies (CAs) obtained with the three experimental conditions (i.e. baseline, CC, and RF) and six binary classification tasks for alpha and beta frequency bands, respectively. For both frequency bands, CC and RF provided statistically significant improvement as compared to baseline in terms of CAs for all the six binary classification tasks. Although RF performed slightly better as compared to the CC method in overall comparison, we obtained a statistically significant difference for only F-W and F-M in  $\alpha$  band and for H-F and F-M in  $\beta$  band among all the six binary tasks. Moreover, for both bands, mixed imagery task pairs (i.e. H-W, F-W) provided maximum separability as compared to the MI task pair (i.e. H-F) and cognitive imagery task pair (i.e. W-M).

Table I and II provides the optimum number of channels (i.e. set of channels resulting in maximum CA) for the

two channel selection methods (i.e. CC, and RF) and six binary classification tasks for  $\alpha$  and  $\beta$  frequency bands, respectively. These results show significant reduction in the number of channels (ranging from 1 to 22) as the number of channels for baseline is 204. Furthermore, the overall number of optimum channels provided by the CC method is lower as compared to the RF method for the majority of subjects.

Table III provides the grand average CA for the three methods over the 15 participants and six binary tasks. For  $\alpha$  band, we obtained the grand average CA ( $\pm$  SD) of 65.32% ( $\pm$  8.51), 80.20% ( $\pm$  8.66), and 81.14% ( $\pm$  8.05) using baseline, CC, and RF methods, respectively. Likewise, for  $\beta$  band, we obtained the grand average CA ( $\pm$  SD) of 59.87% ( $\pm$  7.39), 72.39% ( $\pm$  8.32), and 74.02% ( $\pm$  8.08) using baseline, CC, and RF methods, respectively. Comparing the results of both frequency bands, all three competing methods provided higher CAs for the  $\alpha$  frequency band.

### IV. DISCUSSION & CONCLUSION

The paper has discussed the effect of channel selection and how it has improved the accuracy of the BCI classification tasks. It is also observed, by the tabular presentation of

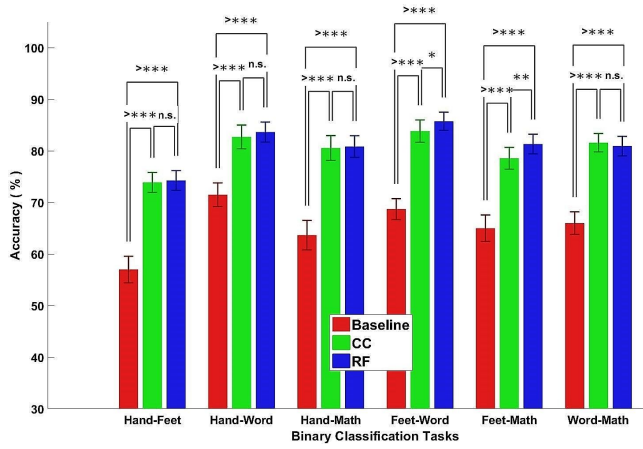


Fig. 3. Mean classification accuracies (CAs) obtained with the three experimental conditions (i.e. baseline, CC, and RF) and six binary classification tasks for  $\alpha$  frequency bands. Error bar represents (SME). \*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$

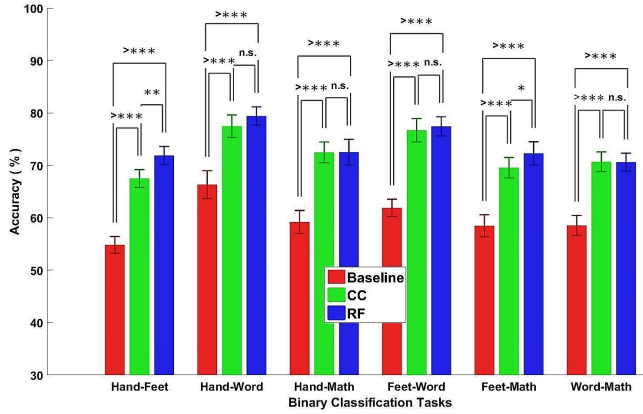


Fig. 4. Mean classification accuracies (CAs) obtained with the three experimental conditions (i.e. baseline, CC, and RF) and six binary classification tasks for  $\beta$  frequency bands. Error bar represents (SME). \*:  $p < 0.05$ , \*\*:  $p < 0.01$ , \*\*\*:  $p < 0.001$

channels contributing towards optimum accuracy, that different methods have different outcomes in terms of accuracy and a total number of channels selected as well as how they vary for every subject. Hence, the channel selection step should be performed during the pre-processing step. Furthermore, it was also observed that there was no statistical significant difference in the classification accuracy between the RF and CC methods using t-test in MI classification. CA calculated in the  $\alpha$  frequency band was better than in the  $\beta$  frequency band in every binary classification task pair. The classification accuracy improved by using the top-ranked channels in both the RF and CC methods. On average the accuracy improved up to 24.22% for the  $\alpha$  band using the RF method. It can thus be concluded that selecting optimal channel will not only increase the accuracy but also reduce the number of channels used dramatically (e.g. in range of 1-22 channels in the above mentioned results), which will have a positive effect on real-time computation.

It is intended to apply other feature selection algorithms

and examine the effects on classification accuracy. The work will also be focused on the method which will provide the most numerically stable computation. As MEG head movement is a major issue, it would be interesting to see the classification accuracy, when the classifier is trained with selected channels and tested on blind session data. In future, the methods will also be tested on the basis of computation cost as well.

## REFERENCES

- [1] L. R. Hochberg, *et al.*, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, p. 164, 2006.
- [2] G. Prasad, *et al.*, "Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: a feasibility study," *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, p. 60, 2010.
- [3] H. Akolkar, *et al.*, "Optimal design and control of a hand exoskeleton for rehabilitation of stroke patients," pp. 1644–1649, 2009.
- [4] J. Farmer, *et al.*, "Effects of voluntary exercise on synaptic plasticity and gene expression in the dentate gyrus of adult male sprague-dawley rats in vivo," *Neuroscience*, vol. 124, no. 1, pp. 71–79, 2004.
- [5] N. Birbaumer and L. G. Cohen, "Brain-computer interfaces: communication and restoration of movement in paralysis," *The Journal of physiology*, vol. 579, no. 3, pp. 621–636, 2007.
- [6] J. R. Wolpaw, *et al.*, "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [7] G. Dornhege, *et al.*, *Toward brain-computer interfacing*. MIT press, 2007.
- [8] H.-L. Halme and L. Parkkonen, "Comparing features for classification of meg responses to motor imagery," *PLoS one*, vol. 11, no. 12, p. e0168766, 2016.
- [9] D. Rathee, *et al.*, "Current source density estimation enhances the performance of motor-imagery-related brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2461–2471, 2017.
- [10] J. Faller, *et al.*, "Non-motor tasks improve adaptive brain-computer interface performance in users with severe motor impairment," *Frontiers in neuroscience*, vol. 8, p. 320, 2014.
- [11] M. Arvaneh, *et al.*, "Optimizing the channel selection and classification accuracy in eeg-based bci," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1865–1873, 2011.
- [12] Y. Wang, *et al.*, "Common spatial pattern method for channel selection in motor imagery based brain-computer interface," in *Engineering in medicine and biology society, 2005. IEEE-EMBS 2005. 27th Annual international conference of the IEEE*, 2006, pp. 5392–5395.
- [13] C. Sannelli, *et al.*, "On optimal channel configurations for smr-based brain-computer interfaces," *Brain topography*, vol. 23, no. 2, pp. 186–193, 2010.
- [14] L. He, *et al.*, "Bhattacharyya bound based channel selection for classification of motor imageries in eeg signals," in *Control and Decision Conference, 2009. CCDC'09. Chinese*. IEEE, 2009, pp. 2353–2356.
- [15] T. Alotaiby, *et al.*, "A review of channel selection algorithms for eeg signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, p. 66, 2015.
- [16] F. Lotte, *et al.*, "A review of classification algorithms for eeg-based brain-computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.
- [17] W. H. Press, "Numerical recipes in fortran: The art of computing," 1992.
- [18] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Aai*, vol. 2, 1992, pp. 129–134.
- [19] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of relief and rrelieff," *Machine learning*, vol. 53, no. 1-2, pp. 23–69, 2003.
- [20] I. Kononenko, "Estimating attributes: analysis and extensions of relief," in *European conference on machine learning*. Springer, 1994, pp. 171–182.